09/829,831

REMARKS

In view of the following discussion, the Applicants submit that none of the claims now pending in the application is anticipated under the provisions of 35 U.S.C. § 102. Thus, the Applicants believe that all of these claims are now in allowable form.

I. REJECTION OF CLAIMS 1, 7-11 AND 17-21 UNDER 35 U.S.C. § 102

Claims 1, 7-11 and 17-21 stand rejected as being anticipated by the Nitta et al. patent (U.S. 4,881,266, hereinafter "Nitta"). The Applicants respectfully traverse the rejection.

The Applicants respectfully direct the Examiner's attention to the fact that Nitta fails to disclose or suggest the novel invention of producing and providing <u>an endpoint signal corresponding to the occurrence of at least one speech endpoint</u> to a speech processing application, <u>along with a speech signal associated with the endpoint signal</u>, for subsequent processing of the associated speech signal, as claimed in Applicants' amended independent claims 1, 11 and 21, from which claims 2-3, 7-10, 12-13 and 17-20 depend.

In contrast, Nitta teaches providing a speech recognizer or "discriminator" (*See, e.g.,* FIG. 3 of Nitta) with a <u>plurality of calculated similarity metrics</u> that each indicate a likelihood of a <u>segment</u> of a speech signal. Thus, Nitta fails to anticipate Applicants' invention.

Specifically, Nitta teaches a speech recognition system that identifies potential endpoints in an input speech signal based on calculated sound power (*i.e.,* the total sound energy emitted by a source per unit time) at various points in the speech signal. In particular, potential endpoints are identified at points in the speech signal where the sound power exceeds (*e.g.,* a starting point) or falls below (*e.g.,* an ending point) a given threshold for a certain duration of time. Portions of the speech signal bounded by starting and ending points are identified as word "intervals", and feature parameters from these word intervals are sampled. The sampled feature parameters are compared to sampled feature points extracted from melcepstral coefficients corresponding to the speech signal, and metrics of this comparison are provided to the discriminator for

09/829,831

sorting and output as a recognition result.

Thus, Nitta teaches a method that, at best, provides <u>similarity measurements</u> <u>relating to pre-segmented portions of a speech signal</u> to a speech recognizer, which produces a recognition result by sorting these measurements. This is <u>not</u> the same as providing the speech recognizer with an <u>endpoint signal</u> (*e.g.*, a binary or continuously generated signal) <u>corresponding to the occurrence of at least one endpoint in a speech</u> <u>signal</u>, *e.g.*, in order to facilitate <u>subsequent</u> signal segmentation and processing by a speech recognition application.   Nitta thus fails to anticipate a method for processing an input speech signal wherein <u>a speech endpoint signal is produced and provided, along</u> <u>with the input speech signal, to a speech processing application for processing of the</u> <u>input speech signal</u>, as positively claimed by the Applicants in claims 1, 11 and 21.

Specifically, Applicants' claims 1, 11 and 21 positively recite:

1.   A method for processing a speech signal comprising:
       extracting prosodic features from a speech signal;
       modeling the prosodic features to identify at least one speech endpoint;
       <u>producing an endpoint signal corresponding to the occurrence of the at least one</u>
<u>speech endpoint;</u> and
       <u>providing the endpoint signal and the speech signal to a speech processing</u>
<u>application</u> to facilitate subsequent processing of the speech signal. (Emphasis added)

11.   Apparatus for processing a speech signal comprising:
       a prosodic feature extractor for extracting prosodic features from the speech
signal;
       a prosodic feature analyzer for modeling the prosodic features to identify at least
one speech endpoint;
       <u>an endpoint signal producer that produces an endpoint signal corresponding to</u>
<u>the occurrence of the at least one speech endpoint;</u> and
       <u>means for providing the endpoint signal and the speech signal to a speech</u>
<u>processing</u> application to facilitate subsequent processing of the speech signal.
(Emphasis added)

21.   An electronic storage medium for storing a program that, when executed by a
processor, causes a system to perform a method for processing a speech signal
comprising:
       extracting prosodic features from a speech signal;
       modeling the prosodic features to identify at least one speech endpoint;
       <u>producing an endpoint signal corresponding to the occurrence of the at least one</u>
<u>speech endpoint;</u> and

09/829,831

<u>providing the endpoint signal and the speech signal to a speech processing application</u> to facilitate subsequent processing of the speech signal. (Emphasis Added)

In one embodiment, the Applicants' invention is directed to a method for applying prosody-based endpointing to a speech signal. Conventional speech processing techniques that are used to provide signals, based on spoken words or commands (*e.g.*, for controlling devices or software programs), typically are characterized by an inability or difficulty in locating suitable speech segments within the spoken input for processing. Typical endpointing techniques identify the completion of a speech segment or utterance by measuring pauses in the given speech signal. However, since spoken language is not typically produced with such explicit indicators, typical endpointing techniques may misinterpret normal fluctuations in the rhythm of speech, such as mid-sentence pauses, to indicate the completion of an utterance. The resultant translation of a spoken command may therefore be fraught with inaccuracies.

The Applicants' invention facilitates the translation of spoken input by extracting and modeling the prosodic features of an input speech signal in order to identify at least one endpoint in the input speech signal. Output is produced in the form of an endpoint signal that represents the occurrence of the identified endpoint in the input speech signal. Both the input speech signal and the generated endpoint signal are then provided to a separate speech recognition application that uses the endpoint signal to facilitate segmentation and subsequent word recognition of the input speech signal. The resultant translated speech thus more accurately reflects the spoken input.

As discussed above, Nitta does not produce or provide an endpoint signal, but rather produces a plurality of similarity measurements relating to sampled segments of an input speech signal. Therefore, the Applicants submit that independent claims 1, 11 and 21 fully satisfy the requirements of 35 U.S.C. §102 and are patentable thereunder.

Dependent claims 7-10 and 17-20 depend from claims 1 and 11, and recite additional features therefore. As such, and for at least the same reasons set forth above, the Applicants submit that claims 7-10 and 17-20 are not anticipated by the teachings of Nitta. Therefore, the Applicants submit that dependent claims 7-10 and 17-20 also fully satisfy the requirements of 35 U.S.C. §102 and are patentable thereunder.

09/829,831

II. ALLOWABLE SUBJECT MATTER

The Applicants thank the Examiner for his comments regarding the allowability of claims 2-6 and 12-16, if rewritten into independent form including all of the limitations of the base claim and any intervening claims. However, in light of the above arguments, the Applicants respectfully submit that claims 1 and 11, from which claims 2-6 and 12-16 respectively depend, are currently in allowable form, and, as such, claims 2-6 and 12-16 are in allowable form as they stand. Nevertheless, the Applicants have included new claims 22 and 23, which respectively present claims 2 and 12 in independent form, as well as new claim 24, which recites an electronic storage medium that stores a program for performing the method recited in claim 22.

III. CONCLUSION

Thus, the Applicants submit that all of the presented claims now fully satisfy the requirements of 35 U.S.C. §102. Consequently, the Applicants believe that all of these claims are presently in condition for allowance. Accordingly, both reconsideration of this application and its swift passage to issue are earnestly solicited.

If, however, the Examiner believes that there are any unresolved issues requiring the issuance of a final action in any of the claims now pending in the application, it is requested that the Examiner telephone <u>Mr. Kin-Wah Tong, Esq.</u> at (732) 530-9404 so that appropriate arrangements can be made for resolving such issues as expeditiously as possible.

Respectfully submitted,

_5/30/06_
Date

Kin-Wah Tong, Reg. No. 39,400
(732) 530- 9404

Patterson & Sheridan, LLP
595 Shrewsbury Avenue
Shrewsbury, New Jersey 07702